

Part 1: Research Proposal

Lab 2: What Makes a Product Successful? - W203 Section 8

Team Herkimer: Rick Chen, Chase Madson, Maria Manna, Jash Sompalli

Jul 19, 2022

Team Herkimer is a data science team for Acme, Inc., the largest used car dealer in the United Kingdom. As proposed herein, this team will perform an explanatory study on vehicle sales to determine which attributes the company should look for to avoid acquiring vehicles with highly-depreciated resale value.

1 Research Question

At what point(s) does a used car depreciate most dramatically in market value price (Y) as a result of its age (X_1) and mileage (X_2)?

To avoid omitted variable bias, we will consider the potential fixed effects of certain static vehicle attributes: engine size (X_3), class (X_4), transmission type (X_5), fuel type (X_6), and mpg (X_7).

2 Data Source

The team acquired data on used car listings that was scraped from the web by Kaggle user *ADITYA* and linked here: **100,000 UK Used Car Data set**

We operationalize the market value price of a used vehicle by the listing price on the advert in pounds sterling (£), stored in the variable **price**. The mileage of a vehicle is reported directly in the variable **mileage**. The age of a vehicle is given indirectly through the vehicle's registration year, reported in the variable **year**.

3 Unit of Observation

Each record in our data represents a unique vehicle appearing in an online used car listing; any vehicles that appear more than once in duplicate listings will be removed, with the most recent listing being favored.

The data was pulled once in 2020 and reflects active listings at the time it was scraped. Vehicles have been limited down to certain manufacturers.